myPEG User Manual

Table of Contents

Introduction

Welcome

myPEG (my Personal Evolutionary Genomics) is a client-server software application that is used for exploring the functional impact of non-synonymous single nucleotide variants (nSNVs). Using myPEG, one can obtain EvoD, PolyPhen2, SIFT, and consensus predictions for the functional impact of nSNVs as well as infer ancestral alleles (using the 46-species UCSC reference peptide alignments – hg19 assembly) for a given coordinate using Maximum Likelihood (ML) or Maximum Parsimony (MP) methods.

EvoD, PolyPhen2, and SIFT predictions as well as related data are pre-computed and stored on the EvoD server (for large numbers of variants, it is recommended to use the Evod server directly).

myPEG is built on top of the MEGA (Molecular Evolutionary Genetics Analysis) software package and is developed using the Delphi programming language.

Citing myPEG

Kumar S, Sanderford M, Gray VE, Ye J, Liu Li.

Evolutionary diagnosis method for variants in personal exomes.

Nature Methods (2012) Sep;9(9):855-6. doi:10.1038/nmeth.2147.

Disclaimer

Although the utmost care has been taken to ensure the correctness of the software, it is provided "as is," without any warranty of any kind. In no event shall the authors or their employers be considered liable for any damages, including, but not limited to, special, consequential, or other damages. The authors specifically disclaim all other warranties, expressed or implied, including, but not limited to, the determination of the suitability of this product for a specific purpose, use or application.

Note that brand and product names (e.g., Windows and Delphi) are trademarks or registered trademarks of their respective holders.

Copyright

Development Team

Sudhir Kumar[1,2] – Project Director

Glen Stecher[1] – myPEG application developer, myPEG and EvoD websites developer

Maxwell Sanderford[1] – EvoD database administrator, EvoD calculation pipeline developer

Koichiro Tamura[3] – myPEG application developer

Dan Peterson[1] **–** myPEG application developer

[1]Center for Evolutionary Medicine and Informatics, Biodesign Institute, Arizona State University (ASU), Tempe, AZ 85287

[2] School of Life Sciences, ASU, Tempe, AZ 85287

[3]Department of Biology, Tokyo Metropolitan University, Hachioji-shi, Tokyo 192-0397, Japan.

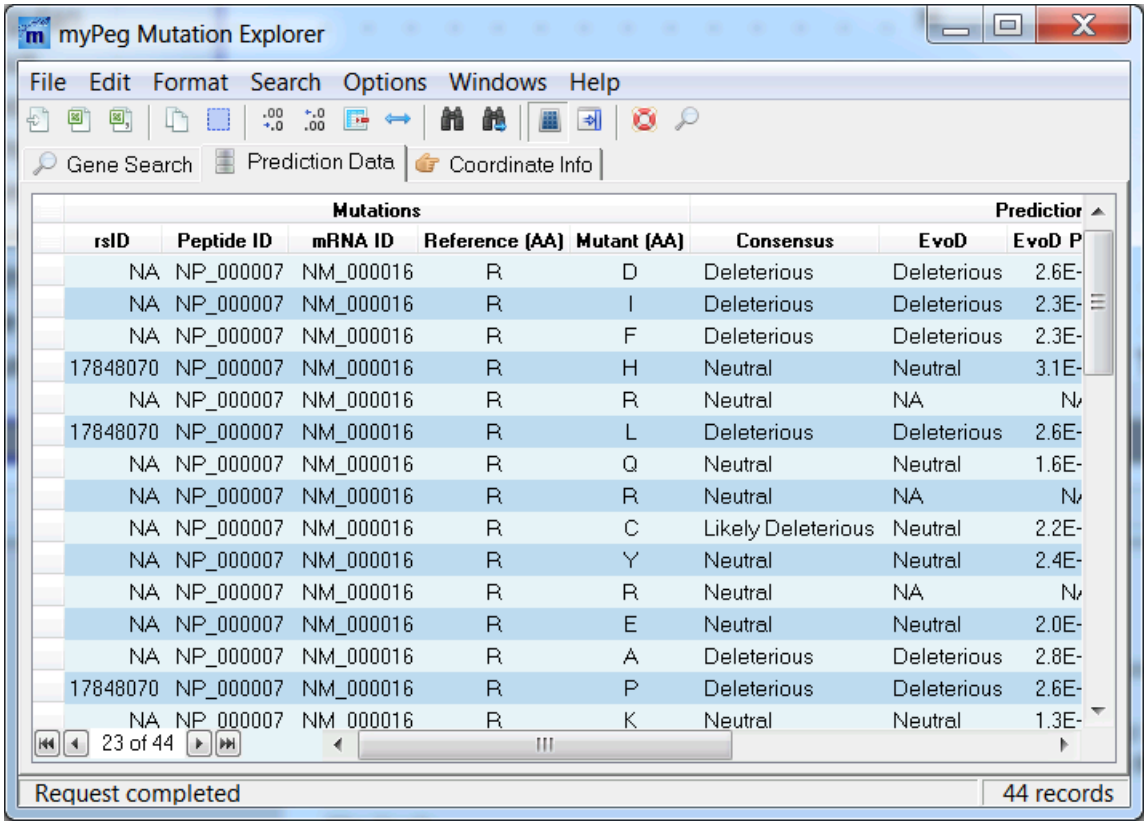
Mutation Explorer

The *Mutation Explorer* window displays data associated with the nSNVs being explored and provides functionality for text searching, sorting, importing, exporting, formatting, gene search, and manual data entry. This window displays three main views, each located on a separate tab:

Gene Search Tab

Prediction Data Tab

Coordinate Info Tab



The actions provided by the Mutation Explorer are divided into several categories and are accessed using the main menu bar or the main tool bar:

File

- Import Query Data From File – load coordinate information form a text file
- Search for a Gene – access the gene search page
- Export Table to Excel File – save all prediction data to an MS Excel file
- Export Table to CSV File – save all prediction data to a Comma-Separated-Values text file
- Exit – Close the application

Edit

- Copy – copy selected values to the system clip-board
- Select All – select all values in the table
- Clear Table – clear all data from the table

Format

- Increase Precision – increase the precision of all numeric values in the table (and also in the *Mutation Detail View* window)
- Decrease Precision - decrease the precision of all numeric values in the table (and also in the *Mutation Detail View* window)
- Resize Columns to Best-fit – resizes all columns in the table to achieve the best fit and optimize the view. Useful when hiding/showing columns and column widths change sub-optimally. ***note: if there are many records in the table (more than several thousand), this operation may take a few moments or more, during which time the window will be unresponsive.

Search

- Find… - text search for values in the table
- Find Next – find the next value matching the search query (search goes to the right and then down to the next row)

Options

- Keep detail view on top – toggle this action on/off to keep the *Mutation Detail View* window from staying in front of other myPEG windows (on by default).
- Show Toolbar – toggle on/off the display of the toolbar (on by default)
- Toggle Auto Column Width – when off (default) a horizontal scroll bar is used to view columns that don't fit in the window. When off, the horizontal scroll bar is removed and all columns are squeezed into view.

Windows

- Detail View Form – show the *Mutation Detail View* window
- Search for a Gene – jump to the *Gene Search* tab in the *Mutation Explorer* window
- Sequence Data Explorer – show the *Sequence Data Explorer* window

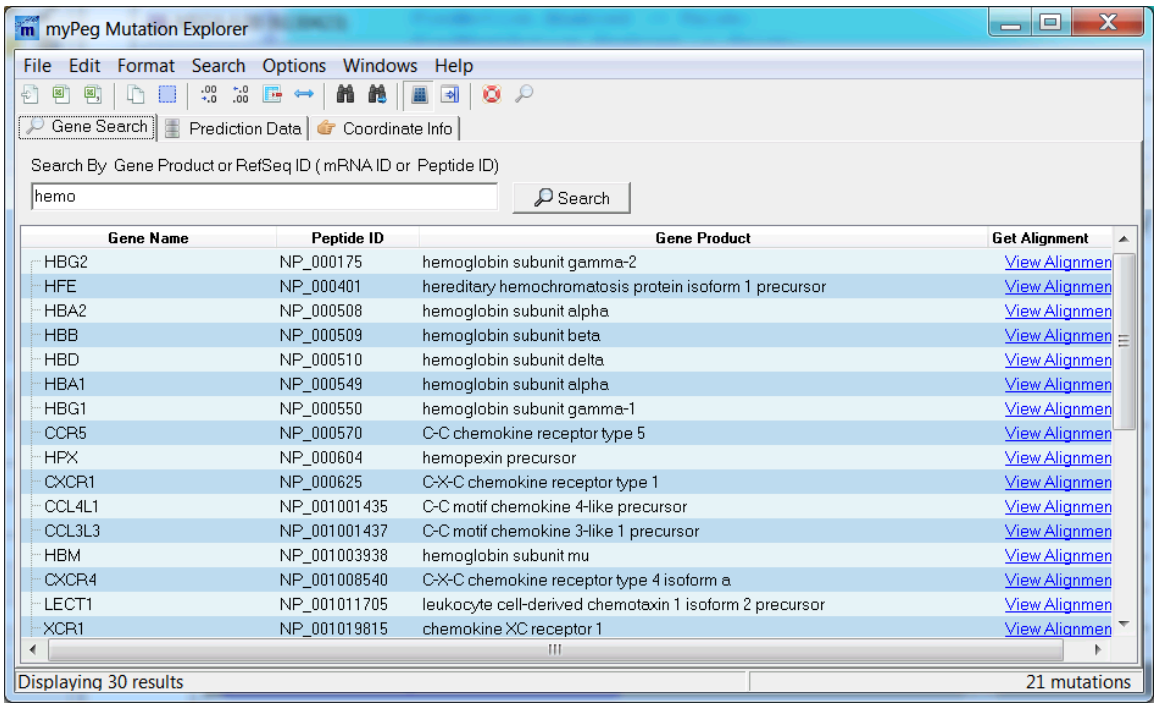
Help

- Contents – Display this help document
- About – show the *About myPEG* window

Gene Search tab

The *Gene Search* tab facilitates searching for genes by keyword (based on gene product) or alternatively by RefSeq identifiers (mRNA ID or Protein ID). Search results (limited to 1000) are displayed in a list view with cursory information and a link for retrieving the 46-species reference protein sequence alignment from the EvoD server. When a sequence alignment is retrieved it is displayed in the Sequence Data Explorer

which can be used to specify the amino acid site and mutant allele for a nSNV of interest.



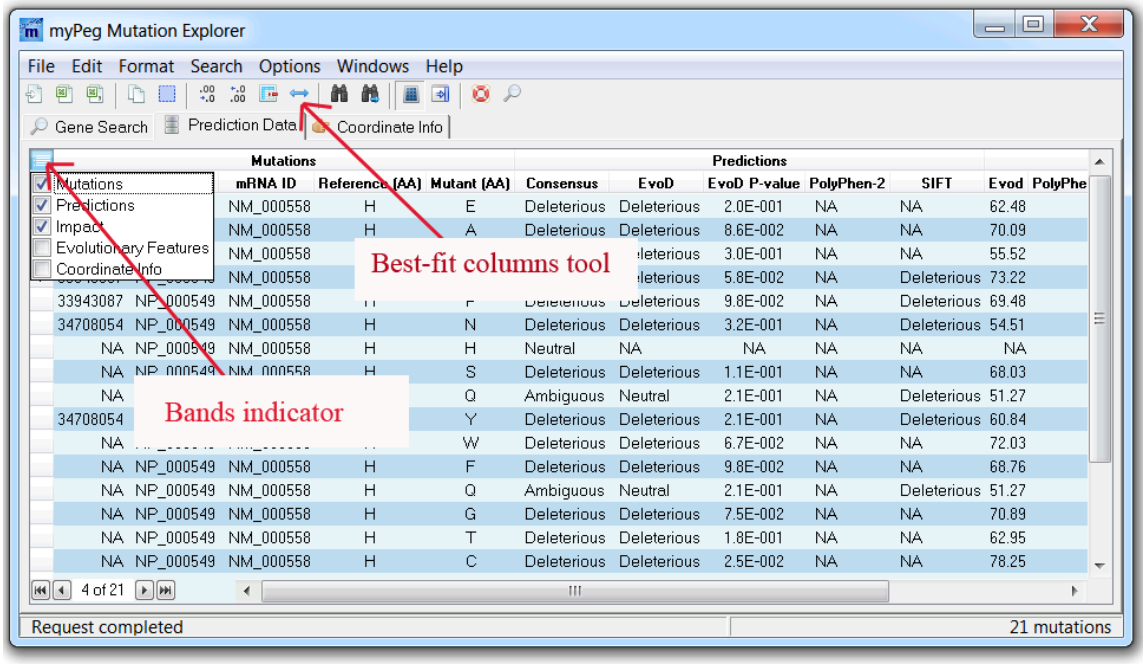
Prediction Data tab

The *Prediction Data* tab displays all prediction data retrieved from the EvoD server in a list view. Complete information for the currently active record is displayed in the Mutation Detail View. Columns of data are banded together into categories:

- Mutations – identifiers as well as mutant and reference alleles are given here.
- Predictions – consensus, EvodD, PolyPhen2, and SIFT predictions are given here.
- *Impact* – the impact scores for EvoD, PolyPhen2, and SIFT predictions are provided along with the Grantham distance and Blosum62 value.
- Evolutionary Features (hidden by default) – substitution rate, position time span, and mutation time span are displayed.
- Coordinate Info (hidden by default) – additional coordinate information is shown here, including chromosome, strand, nucleotide position, amino acid position, wild nucleotide, and mutant nucleotide.

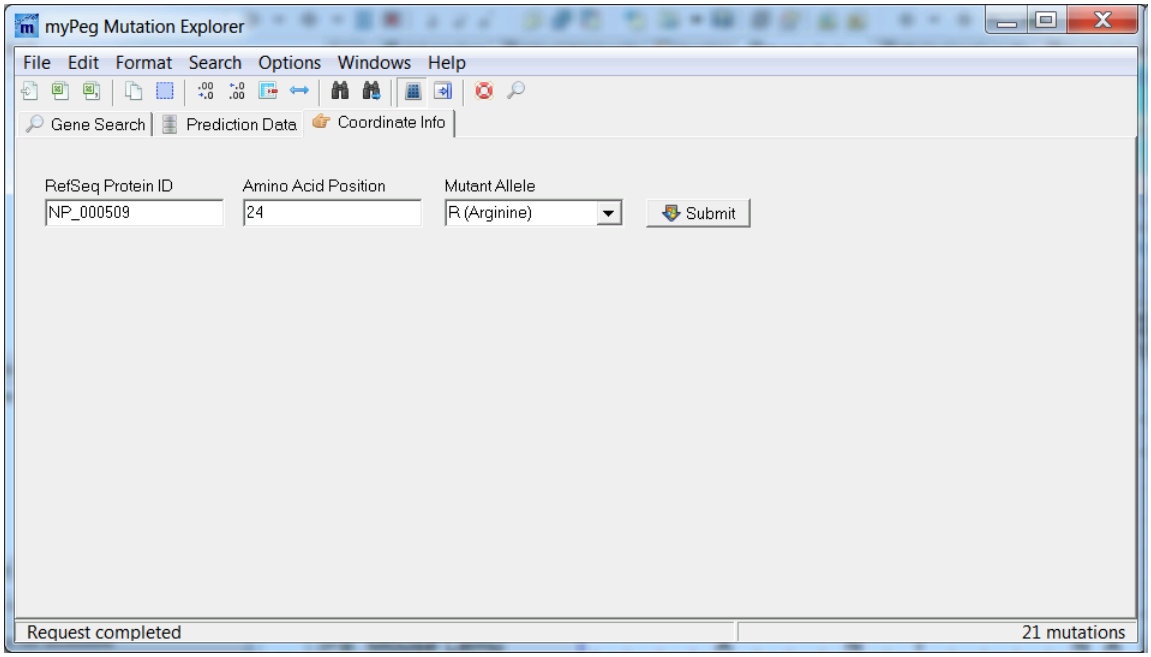
To toggle on/off the display of a given band, click on the indicator button which is located to the far left in the band headers row. A popup menu will appear from which bands can be selected/deselected. Often times when changing the display of bands, column widths will change in undesirable ways. To remedy this, you can execute the *Best-fit Columns* action by clicking *Format->Resize columns to best-fit* or clicking the toolbar button. Alternatively, columns widths can be adjusted by dragging their header edges.

The toolbar and main menu provide access to several actions for importing/exporting data, formatting the view, sorting, text search, and setting view options.



Coordinate Info tab

The *Coordinate Info* tab can be used to input the RefSeq protein id, amino acid position, and mutant allele instead of loading the data from a text file. After required information is entered and the *Submit* button is clicked, myPEG will send query the EvoD server for prediction data. When the data is returned, it is added to the Prediction Data view for further exploration.

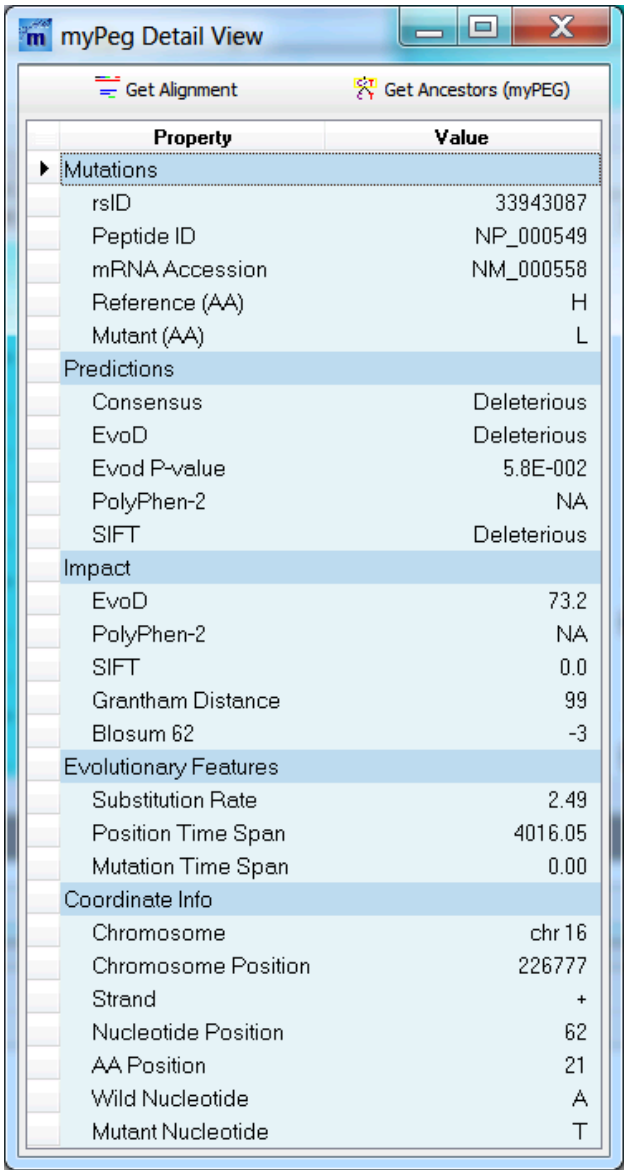


Mutation Detail View

The Mutation Detail View window displays all available information for the currently active record (selected in the Mutation Explorer window). Additionally, this window provides access to the 46-species reference alignment for the given gene as well as the ability to infer ancestral alleles using the Maximum Likelihood (ML) or Maximum Parsimony (MP) methods.

When the *Get Alignment* button is clicked, myPEG will retrieve the 46-species reference alignment from the EvoD server and display it in the Sequence Data Explorer, from which it can be exported or further exploration can be done.

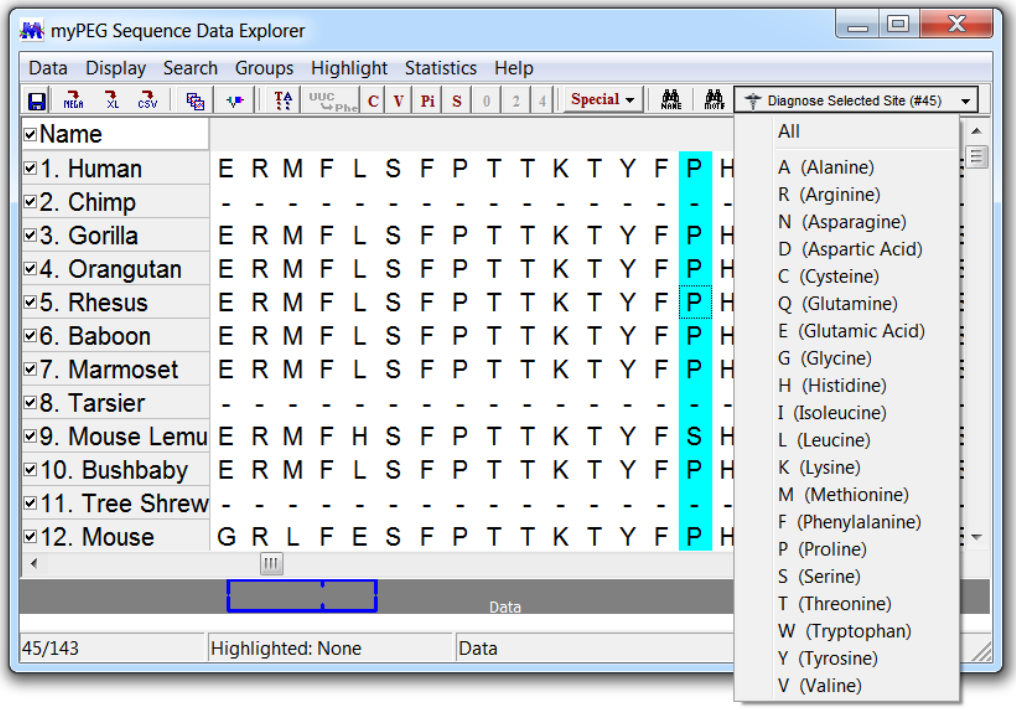
When the *Get Ancestors* button is clicked, the choice of ML and MP methods are presented. After a method is selected the Analysis Preferences Dialog is displayed from which the analysis can be launched with custom settings (e.g. substitution model, distribution of rates, etc…). When the analysis is completed, the reference topology will be displayed in the Tree Explorer along with inferred ancestral alleles for the amino acid site designated earlier.



Sequence Data Explorer

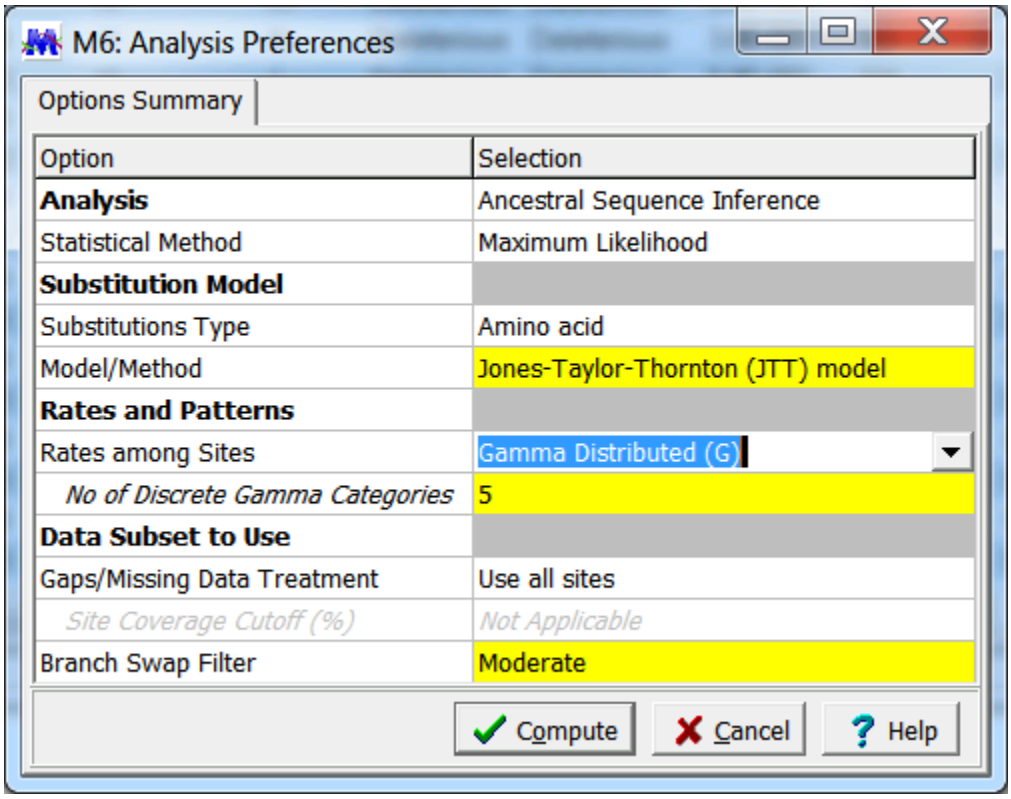
The *Sequence Data Explorer* is used to display the 46-species alignment for a given gene and provides a graphical interface for specifying amino acid position and mutant allele for nSNVs of interest. With an alignment activated, the amino acid position is specified by selecting the site of interest (which will be highlighted). With the site of interest selected, the mutant allele (or all alleles) can be specified from the *Diagnose Selected Site* drop down list. When an allele is selected from the list, myPEG will query the EvoD server and append the returned prediction data to the Mutation Explorer prediction data view.

The *Sequence Data Explorer* window also provides much other functionality such as alignment export and composition based exploration.



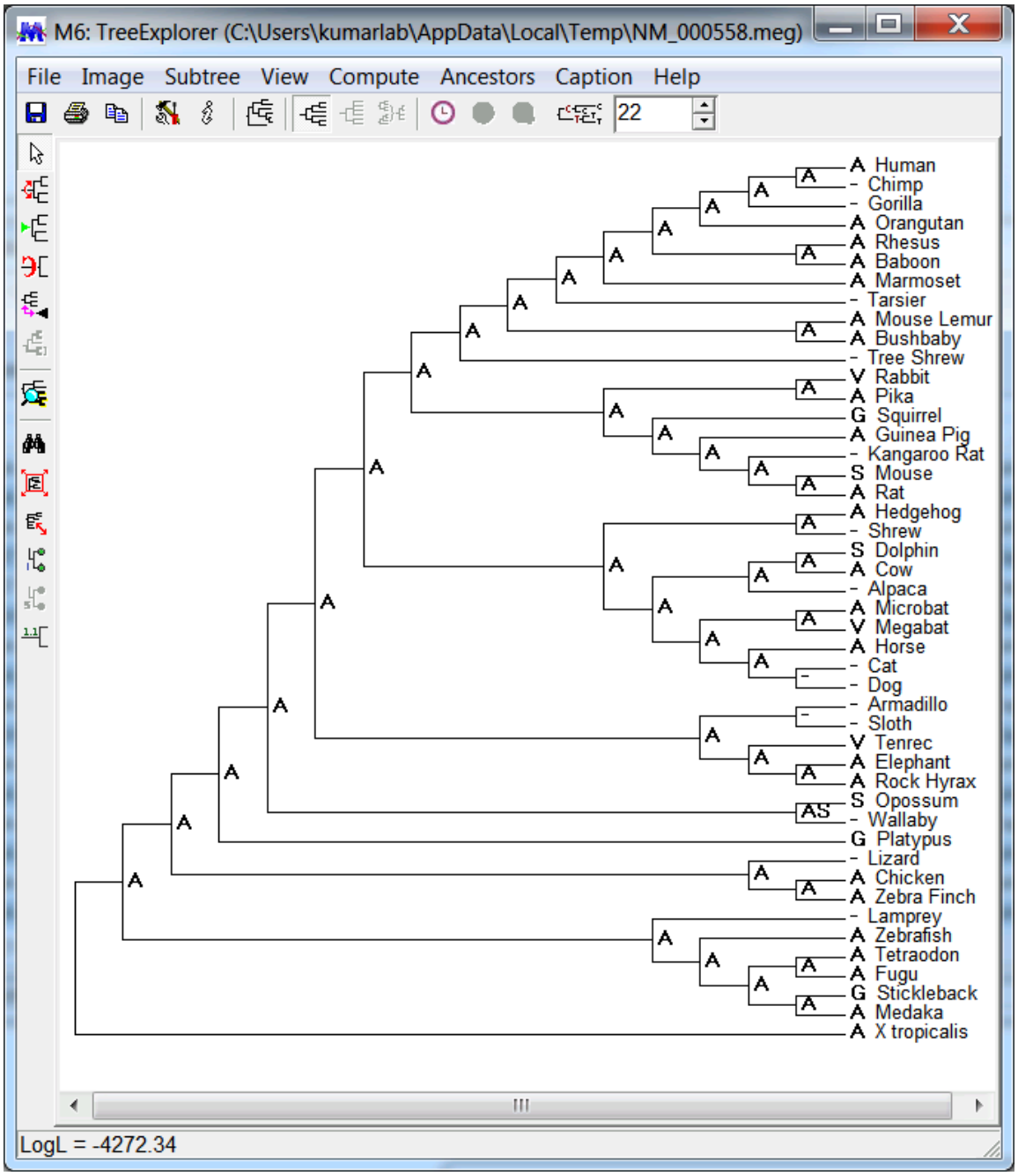
Analysis Preferences Dialog

The *Analysis Preferences Dialog* is used for specifying the substitution model to use as well as the distribution of rates for ML based ancestral sequence inference.



Tree Explorer

The Tree Explorer displays the results of the ancestral sequence inference analyses. When an ancestral sequence inference analysis is complete, the 46-species reference phylogeny is displayed in the Tree Explorer with the inferred ancestral alleles for the nSNV amino acid site shown. The Tree Explorer provides many capabilities which are described in detail in the MEGA5 user manual which can be obtained from www.megasoftware.net/manual.pdf .



Input Data

Overview

In order to retrieve predictions for a given nSNV, myPEG requires three pieces of information:

1. RefSeq protein id (e.g. NP_000082)
2. amino acid position (e.g. 43)
3. mutant allele (e.g. R)

There are three ways to provide this coordinate information to myPEG:

Upload a text file

Manually enter the information

Use the Gene Search and integrated Sequence Data Explorer

Upload a text file with the coordinate information for all nSNVs of interest

Create a text file with coordinate information for all nSNVs to be explored following the format below:

NP_000758	99	E
NP_000761	264	M
NP_000762	144	C
NP_000762	335	W
NP_000773	374	T
NP_000838	71	L
NP_000886	131	H
NP_000887	271	T

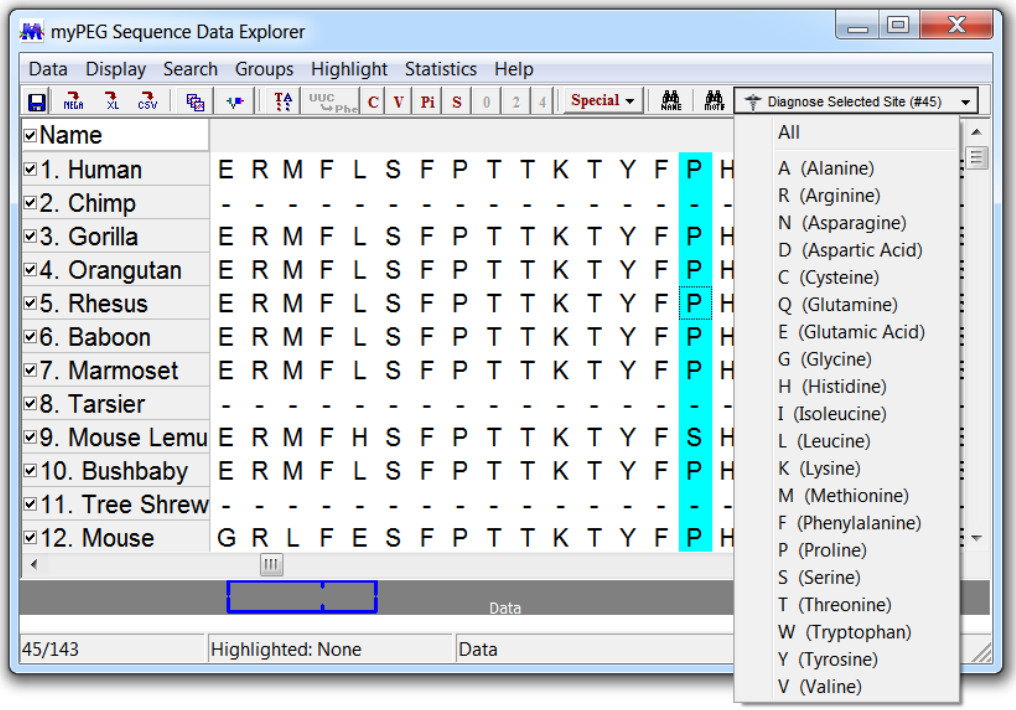
Each line contains coordinate information for one nSNV and each value is separated by white space (*i.e.* spaces or tabs).

In the Mutation Explorer window, select *File->Import Query Data From File* (or click the upload data button) and browse for the newly created text file. myPEG will first validate the format of the coordinate information file and then request prediction information for all specified nSNVs from the EvoD web server. As data is retrieved, the Mutation Explorer window is updated.

The myPEG application has no limit on the number of entries that can be included in the coordinate information file. However, depending on your internet connection speed and the current load on the EvoD server, retrieval of many predictions may take some time (anything less than 5,000 should not be problematic). For situations where myPEG does not perform optimally due to high numbers of nSNVs, the EvoD serverEvoD_Server can be used directly (www.barn.asu.edu/EvoD). The same text file can be uploaded to the EvoD server which will process the file and send you an email for retrieving prediction data once the processing is complete.

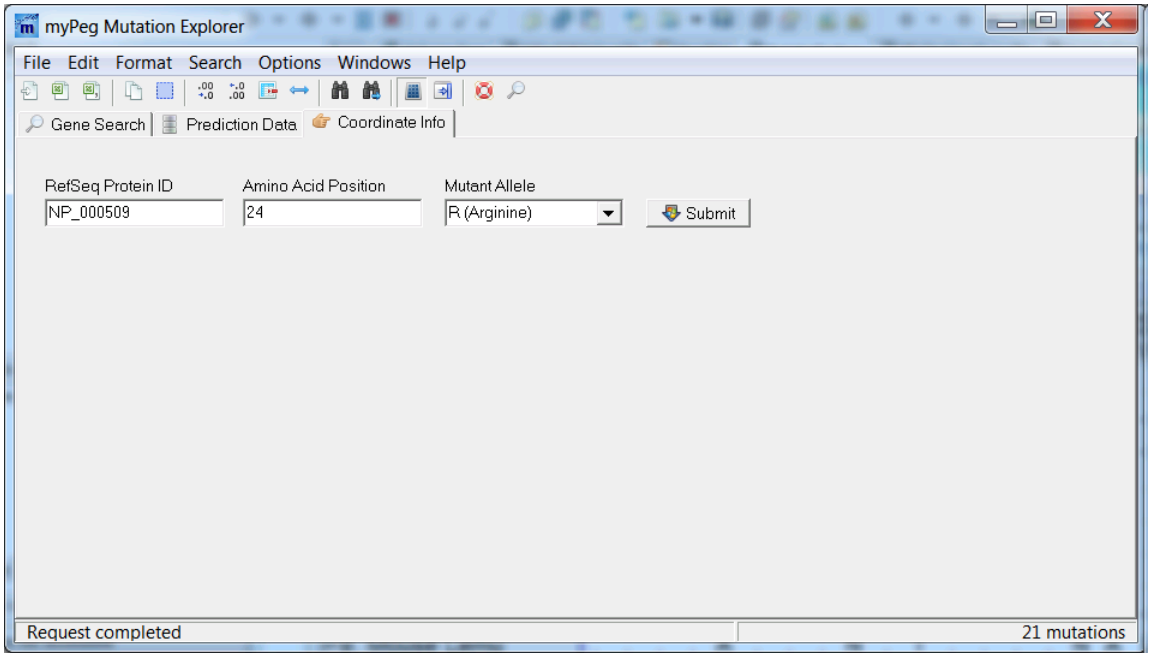
Specify the coordinate information using the integrated Sequence Data Explorer

If a 46-species sequence alignment has been retrieved (see Gene Search) for a given gene, the Sequence Data Explorer window can be used to first navigate to the amino acid site of interest and then specify a mutant allele.



Manually enter the coordinate information using the integrated entry form

In the *Mutation Explorer* window, select the *Coordinate Info* tab. Enter the RefSeq protein id and the amino acid position for the nSNV of interest. Select a mutant allele from the drop down list. Click the submit button and myPEG will send the request to the EvoD server and add the returned data to the *Mutation Explorer* prediction data view.



References

EvoD - Evolutionary Diagnosis

Kumar S, Sanderford M, Gray VE, Ye J, Liu Li.

Evolutionary diagnosis method for variants in personal exomes.

Nature Methods (2012) Sep;9(9):855-6. doi:10.1038/nmeth.2147.

PolyPhen2

Adzhubei IA, Schmidt S, Peshkin L, Ramensky V, Gerasimova A, Bork P, Kondrashov A, Sunyaev S.

A method and server for predicting damaging missense mutations. Nature Methods (2010) 7: 248-249.

SIFT

Pauline C. Ng and Steven Henikoff
SIFT: predicting amino acid changes that affect protein function. Nucl. Acids Res. (2003) 31(13): 3812-3814 doi:10.1093/nar/gkg509

Index